DOCUMENT RESUME

ED 198 486 CS 005 875

AUTHOR Klein, Howard
TITLE Finding Reading Levels with Cloze: A Textbook Dilemma.
PUB DATE [80]
NOTE 21p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; Junior High Schools; Readability; Reading Ability; *Reading Research; *Research Problems; Scoring; *Scoring Formulas; *Textbooks
IDENTIFIERS Bormuth (John R)

ABSTRACT

Current textbooks on reading recommend six scoring systems to be used with cloze tests to find material of suitable difficulty for instruction. These six scoring systems, when applied to a single data source--300 cloze scores obtained from ninth grade students tested with ninth grade content materials--produced varied placements which at present cannot be converted from one to another. Obviously all six scoring systems, differing as they do, cannot make correct placements. Perhaps the greatest disservice of textbook authors, aside from not being up to date and comprehensive in their reporting on cloze procedures, is that they have neglected John Bormuth's evolution in thinking about the rationale upon which many cloze scoring systems have been built. Bormuth particularly singles out criterion scores used with multiple choice tests. If teachers choose to follow the recommendations of textbooks on reading, these textbooks will have a profound effect in the classroom--but there is no basis for knowing whether these effects will be good or bad. (RL)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED198486

# Finding Reading Levels with Cloze:

## A Textbook Dilemma

Dr. Howard Klein

Department of Curriculum Studies

College of Education

University of Saskatchewan

Saskatoon, Saskatchewan

S7N 0W0

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Howard Klein

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS008825

Almost three decades have passed since Taylor (1953) introduced cloze. In the interim, the methodology has had a penetrating influence on literacy research, and today's reading textbooks, almost without exception, devote several pages to a description of the procedure and a discussion of one or more of its multitudinous uses.

## TEXTBOOKS AND TEACHERS

Teachers seeking information about how cloze matches reader to print material will, in all likelihood, refer to a textbook (Mathewson, 1979). Line (1971) is convinced teachers prefer

> carefully prepared and temptingly presented packages summarizing and evaluating research findings that are sufficiently established for their practical implications to be clear (p. 424).

He, therefore, concludes teachers tend to spurn raw research found in monographs and journals. Should teachers put into practice recommendations made in textbooks, when textbooks designed for courses are not the sole source of information in those courses?

A textbook is a standard work of basic knowledge written by experts to support a program of instruction in a particular branch of study such as reading. Textbooks provide indirect experiences in large and well organized amounts. This is accomplished through a synthesization of extant knowledge, which appears in a selected and ordered way (Hilton, 1969). Of necessity, perhaps, every textbook must be limited by selection. That process determines what will be included and how much can be omitted. Specifically, authors decide how much of the literature will be interpreted. Relevant to this paper is the omnipresent scoring

system to be used with cloze for matching child to book. In general, textbooks present in detail the construction of a cloze test, identify the scoring criteria they recommend, and indicate its source, but no rationale is provided for the scoring system. Missing too, is supportint research demonstrating its efficacy.

Very strange, indeed, is the reluctance of most authors to acknowledge the existence of other scoring systems. Cloze, then, is a superb instance of how selection and, particularly, omission work in current reading textbooks.

## CLOZE SCORING SYSTEMS

Most reading textbooks published since 1976 discuss the merits of cloze as an assessment device to place children in reading material of appropriate difficulty. Most of the textbook authors present scoring criteria to determine three reading levels: independent, instruction and frustration. In all, six distinct scoring systems have been adapted by textbook authors. Using authors' citations, the scoring criteria they recommend can be traced to four studies (Bormuth, 1967; 1968b; Rankin and Culhane, 1969), with oral reading tests (Bormuth, 1968a), or with informal reading inventories (Jones and Pikulski, 1974). To compare cloze to another type test, the strategy used by the researchers was to make two tests from the same passage, for example, cloze and multiple-choice questions, thus holding content constant. In order to relate cloze to multiple-choice, Bormuth (1967) and later Rankin and Culhane (1969) selected criterion scores of 75 percent for instruction and 90 percent for independence as benchmarks for multiple-choice tests. The researchers were satisfied with those standards, because they had been suggested by respected experts and enjoyed a long history of wide

use. Although Bormuth (1969) accepted those standards, two years later he (Bormuth, 1971) had second thoughts and,for all intent and purpose, repudiated the 75-90 criteria for want of scientific evidence that they are useful (see also Bormuth, 1975). Interestingly enough, textbooks, by and large, do not discuss Bormuth's evolution which resulted in new scoring criteria.

Table one reveals the scoring systems developed in the field have great variation. In turn, twelve textbook adaptations maintain those differences. Almost without exception, the authors of those textbooks fail to explain why they selected a particular scoring system; nor do they give reasons for any deviations from the original when they occur.

INSERT TABLE ONE ABOUT HERE

Although there are six scoring systems, only four scores, ranging from 29 to 43 percent determine the upper limits marking frustration, and another four scores, with a range of 51 to 61 percent are the lower limits for independence. The scores falling between these criteria form bands of 10 to 20 percentage points denoting the instructional level. Only seven scores--44 to 50 percent are common to all systems. The question in need of answering is how would a set of cloze scores be affected if all six textbook scoring systems are applied to the scores?

## PROCEDURES

### The sample

In order to determine how scoring systems place students, a large number of cloze scores were required and these came from a data bank. Ultimately 300 scores were used from that bank for this investigation. Those scores came from ninth grade boys and girls who attended five of seven public high schools located in a western Canadian city. Each

English department in each of the five participating schools identified classes representative of the grade nine students enrolled in the school.

The original sample consisted of over 450 students from the 18 identified classes. The students were tested in their classrooms and were randomly divided into nine groups by distributing pre-mixed booklets according to a set plan for all classrooms. The booklets had been sorted using a table of random numbers, and distribution from the top of the pack began with the first student in the first row on the right. The procedure gave every student an opportunity to write any one of the nine cloze tests in the battery.

THE PASSAGES

Nine cloze tests were prepared from nine 500 word passages. The passages had been systematically winnowed from 30 ninth grade textbooks available for instruction in the province, ten in each of three subject areas: English literature, social studies and science. Twenty-four randomly selected samples of approximately 100 words were taken from each book. Three measurements of each passage's readability were determined with Fry, Dale-Chall and SMOG formulas. A mean grade was found for each of the 30 textbooks with each formula, and the books were ranked within subject areas from high to low difficulty. For each subject area three textbooks were identified representing the highest, the average and the lowest readability levels found for the ten books.

The Fry Formula was selected to identify cloze test passages, because its measurements consistently fell between Dale-Chall and SMOG and correlated .95 with the former and .91 with the latter. From each book, the selected passage was equated to the mean difficulty of the book. Cloze tests, with an every fifth word deletion rate, were prepared. Prior to testing, each class was shown how to take the cloze test using standard

instructions and sample material. Exact replacements were scored correct.

The 300 test scores for this study are from six of the nine tests: three social studies, two science and one English literature. Statistical analysis of the passages indicated differences existed among the nine passages, but the Newman-Keuls Multiple Range Test of Differences showed six of the passages are statistically equivalent with a mean score or 39.45 and a passage range from 37.96 to 42.32. For the purpose of this investigation, all scores were ranked on a compilation sheet, and all scores were then classified six times according to the criteria identified by each of the six scoring systems. A total number for each cell was recorded, and, then, that number was translated into a percent.

FINDINGS

Although the six scoring systems designated those falling within instruction in bands of from 10 to 20 percentage points, the only scores for which there is concurrence are 44 to 50 percent, and they account for 23 percent of the sample across systems. Individual scoring systems identified from 29 to 82 percent of the sample as instructional, when reading ninth grade materials, found to be of comparable difficulty using cloze, despite measured readability differences.

Four textbooks accept a criterion score of 40 percent designating instruction and five textbooks choose 44 percent. Figure one shows that a difference of four percentage points changes placements of one in five

INSERT FIGURE ONE ABOUT HERE

individuals from instruction to frustration and visa versa. This may be attributed to the mean falling close to these criterion scores. A second analysis of scores from 377 fifth graders reveals similar fluctuations. The mean score is approximately 47 percent, and individual scoring systems

change placements of 25 percent of the sample between instruction and independence. In contrast to this study, the two criterion scores,40 and 44 percent,affect only eight percent of that sample.

## DISCUSSION

Current reading textbooks recommend six scoring systems to be used with cloze tests to find material of suitable difficulty for instruction. The six scoring systems,when applied to a single data source,produce varied placements which, at this time, cannot, like fahrenheit and celsius temperature scales, be converted one to another. Obviously, all six scoring systems, differing as they do, cannot make correct placements.

Perhaps it is too early to know which system is correct, but textbooks writers could partially solve the problem inherent in disagreement by following the example of Guszak and Hofman (1980) who opted for Rankin and Culhane's (1969) system but acknowledge, in a single sentence, the existence of different criteria from Bormuth (1967). Stauffer et al. (1978:157), whose system only identified ten percent of this study's sample in frustration, may well have some doubts about its utility, commenting, "Other authors suggest more stringent criteria." They explain their scores are starting points and can be adjusted "based upon practical experience with a particular population of readers or with a particular type of reading material." However, the authors offer no guidelines for making adjustments and they make no mention of other scoring systems.

## OTHER PROBLEMS

Decisions governing textbook adaptations of scoring systems become more mysterious when reading two textbooks authored by the same individuals aimed at two audiences--one written for elementary and the other for secondary teachers. For some inexplicable reason, Karlin (1977; 1980)

and Burns and Roe (1980; Roe, Stoodt and Burns, 1977) selected two scoring systems, one for each textbook. In the later book, they ignore the other system. Therefore, all four books perpetuate the impression that the recommended scoring system is without an alternative. What scholarly intuition determined one system was better than another for two different levels when the cited research used intermediate students?

Estes and Vaughan (1978) contribute to misunderstanding. In a sense, they indict the scholarship of all reading textbook authors who state or imply their scoring systems can be used to place individuals. Estes and Vaughan (1978:31) advise the user of their system "to compute cloze scores and average them for a class." They argue

> To our knowledge, no research had found cloze to be valid for an individual's score. Hence it is the group's average score that is compared to the criteria (page 31).

Is their interpretation of the literature correct? If they are right, then every other author must be wrong. While there appears to be little concurrence on scoring criteria and no evidence that placements with cloze are effective in a classroom of children, Bormuth (1975) does recommend placing individuals with cloze

> With a single student who has taken one cloze test over each of several sets of materials, we set the highest value on the material on which his or her test scores most closely approaches the criterion score... (page 81)

On the other hand he does recommend that cloze be used to select a text, though he (page 82) cautions "we would not choose a single text

because great differences among individuals make any single text poorly suited to a large number of students." He does advise

> To select such a text, choose material on which the mean cloze score falls somewhat above the criterion score, assuring that the material is partially comprehensible to nearly everyone. (page 82).

The literature demonstrates that the IRI, with a much longer history, is also beset with a scoring criteria controversy (Beldin, 1970; Powell, 1971). Homan (1978) concluded that the scoring criteria used for the informal can make very significant differences in the accuracy of placements. Walters (1975) noted that the existence of high correlation coefficients between cloze and the informal reading inventory are inadequate criteria on which to accept cloze as a valid technique for determining instructional levels of pupils. He compared cloze and an informal reading inventory placements and found 31 percent of the children to have been assigned identical instructional levels with the two measures. Pollock (1975) compared cloze to the informal reading inventory and reported only 29 percent of the subjects were assigned the same reading level.

## BORMUTH'S EVOLUTION

Even though there has been enough lead time, current textbook authors have ignored Bormuth's (1971; 1975) evolution in thinking about literacy, readability and instructional placement. Donnelly (1980), alone, explores new criteria (Bormuth, 1969), but does so through Earle's (1976) interpretation. In 1969, Bormuth had not yet disavowed the basis for his earlier scoring systems.

Even as Bormuth was disseminating information about using cloze to match materials to the reader, his theoretical constructs and resultant research were undergoing a transformation, moving beyond the systems he had been publicizing which were beginning to reach teachers through textbooks. Perhaps the greatest disservice of textbook authors is that they have neglected his questioning the rationale upon which many scoring systems have been built. He particulary singles out criterion scores--75 and 90 percent--used with multiple-choice tests.

Bormuth (1967) and Rankin and Culhane (1969) matched cloze to multiple-choice scores. Up to that time, there was uncertainty about what a cloze percent score meant, but benchmarks for multiple-choice testing were available. Bormuth (1967) wrote

> The reasoning underlying this study is that multiple-choice comprehension tests have been used for many years and have a widely known frame of reference accepted in both readability research... and in classroom practice... (page 202).

He insisted that multiple-choice scores falling between criteria of 75 and 90 percent for specific material are suitable for use in supervised instruction. At that time,he accepted expert opinion and usage as substantive evidence of the scores' utility.

Some say Development of Standards of Readability: Toward a Rational Criterion of Passage Performance (Bormuth, 1971) is "the most definitive research done on the topic (Binkley, 1980). However, defensible that assessment may be, textbooks do not cite it. Bormuth (1971) argues criterion scores used to interpret students' scores in the past have been based on shaky reasoning and scanty evidence. Bormuth decided the re-

nowned experts never provided evidence that the criteria they proposed were useful or explained their reasoning for selecting those particular scores representing the "considered judgments of what constitutes the most desirable performance levels..." (Bormuth, 1975:62). Intuitive judgments, Bormuth observes, could lead to extraordinary discoveries or to false, but widely accepted beliefs such as the need to bleed devils out of the sick people or the cause of thunder is warring gods. Clearly, the problem with scores--75 and 90 percent--is that common sense and intuition are not enough. Those scores have not been examined critically and verified or refuted scientifically.

Bormuth (1971) chose an entirely different tack, developing a new method for establishing cloze criteria which produced a new scoring system popularized in 1975. Bormuth's new criterion marking instruction is 35 percent. Individuals scoring below that level would find the material too difficult for learning. However, the material would be easy for those who score above 50 percent. Bormuth (1971) is cautiously optimistic about what he has achieved:

> While much work remains to be done before fully acceptable criterion scores are obtained, the criterion scores obtained appeared to be somewhat useful. (page ix)

Nevertheless, he is conviced,

> The criterion scores presented here are better by several orders of magnitude than those presently in use. (page 148)

This analysis of cloze scoring systems reveals the dilemma caused by reading textbooks. If teachers choose to follow their recommendations, they can have a profound effect in the classroom, but there is no basis

for knowing if what happens is good or bad. In most cases, the adoption of a scoring system may have been at the intuitive level with the possible exception of Stauffer et al. (1978) where family was involved. Teachers should be alarmed that authors of textbooks fail to overcome time lag. Granted, they are aware of Bormuth's change of direction; nevertheless, books with 1980 copyright dates make no concession to research or deal with the questions he has raised. Ostensibly textbook writers do not accept his claim that his new criteria are vastly superior to what already exists. Why?

## CONCLUSION

This paper analyzes and discusses six cloze scoring systems identified in twelve current reading textbooks. The conclusion is that six scoring systems are five too many. Probably they are in the textbook literature because of inherent problems in reading textbooks.

Is cloze the only textbook dilemma?

---

The writer gratefully achnowledges Mother M. Berchmanna (deceased) and Don Bewell whose data were used in this paper.

TABLE 1

RECOMMENDED CLOZE PERCENT SCORES FOR IDENTIFYING THREE READING LEVELS

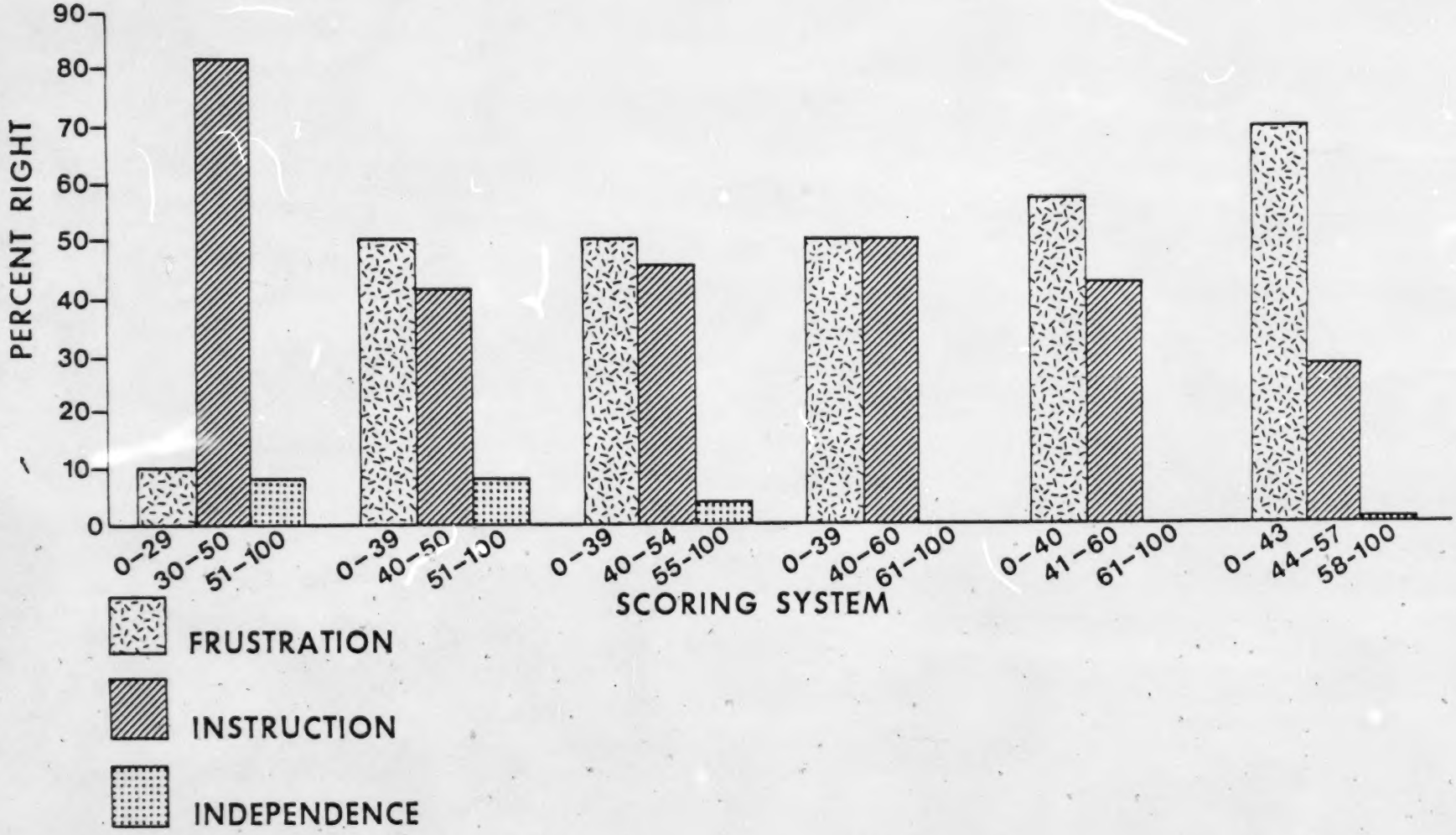
| Textbook's Recommended Standards | | | Textbook | Cited Study | Researcher's Recommended Standards | | |
|---|---|---|---|---|---|---|---|
| Frus | Inst | Ind | | | Frus | Inst | Ind |
| 0-29 | 30-50 | 51-100 | Stauffer, Abrams & Pikulski (1980) | Jones & Pikulski (1974) | 0-20<br>0-24 | 21-49<br>25-42 | 50-100<br>43-100 |
| 0-39 | 40-50 | 51-100 | Burns & Roe (1980)<br>Zintz (1977) | Bormuth (1967) | 0-37 | 38-39 | 50-100 |
| 0-39 | 40-54 | 55-100 | Karlin (1980) | Bormuth (1968a)(Word Recog) | 0-33 | 34-53 | 54-100 |
| 0-39 | 40-60 | 61-100 | Estes & Vaughan (1978)[2] | Bormuth (1968b) | | | |
| 0-40 | 41-60 | 61-100 | Harris & Smith (1980)<br>Rupley & Blair (1978) | Rankin & Culhane (1969) | 0-40 | 41-60 | 61-100 |
| 0-43 | 44-57 | 58-100 | Ekwall (1976); Lapp & Flood (1978); Roe, Stoodt & Burns (1978); Otto, Rude & Speigel (1979); Smith & Robinson (1980) | Bormuth (1968a) Comp | 0-43 | 44-56 | 57-100 |

[1]Karlin appears to have adapted Bormuth's words recognition criteria

[2]Estes & Vaughan do not cite Rankin & Culhane, but their scoring system most approximates that system. They cite Bormuth (1968b).

FIGURE 1

## Reading Placements Based on 300 Cloze Scores As Determined by Six Scoring Systems



# B I B L I O G R A P H Y


Beldin, H. O. Informal reading testing: historical review and review of the research, in W. K. Durr Ed. Reading Difficulties: Diagnosis, Correction, and Remediation. Newark, Deleware: International Reading Association, 1970, 67-84.

Binkley, J. R. Personal Communication, November 19, 1980.

Bormuth, J. R. Comparable cloze and multiple-choice comprehension test, Journal of Reading, 1967, 10, 291-299.

____ Cloze test readability: criterion reference scores, Journal of Educational Measurement, 1968a, 5, 189-196.

____ The cloze readability procedure, Elementary English, 1968b,45, 492-436.

____ Development of readability analysis. Final Report, Project No. 7-0052, Contract No. OEC 3-7-070052-0326. Bureau of Research, Office of Education. U. S. Department of Health, Education, and Welfare, 1969.

____ The cloze procedure: literacy in the classroom, in William D. Page. Ed. Help for the Reading Teacher: New Directions in Research. National Conference on Research in English, Urbana, Illinois: National Council of Teachers of English, 1975, 60-90.

____ Development of Standards of Readability: toward a rational criterion of passage performance. U. S. O. E. Technical Report No. 9-0237 U. S. Office of Education, 1971.

Burns, P. C. & Roe, B. D. Teaching Reading in Today's Schools (2nd ed.). Chicago, Illinois: Rand McNally, 1980.

Donnelly, M. Reading in content areas. in L. Estil Alexander, Ed. Teaching Peading. Boston, Massachusetts: Little, Brown and Company, 1979, 222-252.

Earle, R. A. Teaching Reading and Mathematics, Newark, Delaware: International Reading Association, 1976.

Ekwall, E. Diagnosis and Remediation of the Disabled Reader. Boston, Massachusetts: Allyn and Bacon, 1976.

Estes, T. H. & Vaughan, J. L. Reading and Learning in the Content Classroom. Boston, Massachusetts: Allyn and Bacon, 1976.

Guszak, F. J. & Hoffman, J. V. Comprehension skills in Pose Lamb and Richard Arnold Eds. Teaching Reading: Foundations and Strategies 2nd ed. Belmont, California: Wadsworth Publishing Company, 1980, 309-340.

Harris, L. A. & Smith, C. B. Reading Instruction: Diagnostic Teaching The Classroom 3rd ed., New York, New York: Holt, Rinehart and Winston, 1980.

Hilton, E. Textbooks, in Robert L. Elib, Ed. Encyclopedia of Educational Research, 4th ed., Toronto, Ontario: Collier-MacMillan Limited, 1969, 1470-1476.

Homan, S. L. Determining instructional reading level: an investigation of the relationship between standard cloze tests, multiple-choice cloze tests, and the informal reading inventory (doctoral dissertation, University of Florida, 1978). Dissertation Abstracts International, 1979, 39 7134A (University Microfilm No. 7913286).

Jones, M. B. & Pikulski, E. D. Cloze for the classroom. Journal of Reading, 1974, 17, 432-438.

Karlin, R. Teaching Elementary Reading, 3rd ed., New York, New York: Harcourt, Brace, Jovanovich, 1980.

____ Teaching Reading in High School: Improving Reading in Content Areas, 3rd ed., Indianapolis, Indiana: Bobbs-Merrill Educational Publishing, 1977.

Lapp, D. & Flood, J. Teaching Reading to Every Child. New York, New York: MacMillan, 1978.

Line, M. B. The information uses and needs of social scientists: an overview of INFROSS, ASLIB, 1971, 23, 412-434.

Mathewson, H. J. Information seeking behaviors and attitude to information among educational practitioners, Doctoral Dissertaion, University of British Columbia, Vancouver, B.C., 1979.

Otto, W. Rude, R. & Speigel, D. L. How to Teach Reading. Reading, Massachusetts: Addison-Wesley, 1979.

Pollock, D. H. The use of cloze procedure to determine reading instructional levels (doctoral dissertation, Lehigh University, 1974). Dissertation Abstracts International, 1974, 35, 18794-1880A (University Microfilm No. 74-21, 434).

Powell, W. R. Validity of the IRI reading levels. Elementary English, 1971, 48, 637-642.

Rankin, E. & Culhane, J. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.

Roe, B. D. Stoodt, B. D. & Burns, P. C. Reading Instruction in the Secondary School. Chicago, Illinois: Rand McNally, 1978.

Rupley, W. & Blair, T. R. Reading Diagnosis and Remediation: A Primer for classroom and Clinic. Chicago, Illinois: Rank McNally, 1979.

Smith, N. B. & Robinson, H. A. Reading Instruction for Today's Children, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, 1980.

Stauffer, R. G., Abrams, J. C. & Pikulski, J. J. Diagnosis, Correction and Prevention of Reading Disabilities. New York, New York: Harper and Row, 1978.

Taylor, W. L. Cloze procedures: A new tool for measuring readability. Journalism Quarterly, Fall, 1953, 30, 415-433.

Walter, R. B. An investigation into the relationships between cloze test scores and informal reading inventory scores of fifth grade pupils (doctoral dissertation, Ball State University, 1973). Dissertation Abstracts International, 1973, 34, 2941A-2942A (University Microfilm No. 73-25, 515).

Zintz, M. V. Corrective Reading, 3rd ed., Dubuque, Iowa: William C. Brown, 1977.